

# FULL INFORMATION ESTIMATIONS OF A SYSTEM OF SIMULTANEOUS EQUATIONS WITH ERROR COMPONENT STRUCTURE

PIETRO BALESTRA  
JAYALAKSHMI VARADHARAJAN-KRISHNAKUMAR  
*University of Geneva*

In this paper we develop full information methods for estimating the parameters of a system of simultaneous equations with error component structure and establish relationships between the various structural estimators.

## 1. INTRODUCTION

The error component model is one of the earliest econometric models developed to enable the use of pooled time-series and cross-section data. Studies like those of Balestra and Nerlove [5], Wallace and Hussain [20], Amemiya [1] are just a few references that deal with single-equation error component models. Avery [2] went a step further by combining error components and seemingly unrelated equations, and Magnus [12] offered a complete analysis of estimation by maximum likelihood of multivariate error component models, linear and nonlinear, under various assumptions on the errors. Recently, the error component structure was extended to a system of simultaneous equations by Baltagi [7], Varadharajan [19], and Prucha [18]. In this paper we develop full-information methods for estimating the parameters of a system of simultaneous equations with error components structure and establish relationships between the various structural estimators available. Prucha's article, which appeared as the final version of this paper was submitted for printing, also deals with the full information maximum likelihood (FIML) estimation of a system of simultaneous equations with error component structure. He follows an approach similar to that of Hendry [11] to generate a class of instrumental variable (IV) estimators based on the IV representation of the normal equations of the FIML procedure. We follow a different approach, similar to that of Pollock [16], and derive the limiting distributions of both the coefficient estimators and covariance estimators of the FIML method, whereas Prucha gives the limiting distribution of the coefficient estimators only.

Our paper is organized as follows. Section 1 presents the model and spec-

ifies the notations. In Section 2, maximum likelihood estimation of the reduced form is described and the limiting distribution of the resulting estimators is shown to be the same as that of covariance estimators. In Section 3, alternative specifications of 2SLS and 3SLS, namely “generalized” 2SLS and “generalized” 3SLS, are presented. It is also shown that the “generalized” 2SLS and the indirect least squares (ILS) estimators of the coefficients of a just-identified structural equation have the same limiting distribution and that when the whole system is just-identified, “generalized” 3SLS and “generalized” 2SLS estimators are identical and are both asymptotically equivalent to ILS estimators. Section 4 derives the full information maximum likelihood (FIML) estimates of structural parameters and proves the asymptotic equivalence of the FIML estimators and the “generalized” 3SLS estimators. Finally, conclusions are drawn in Section 5.

## The Model

We consider a system of  $M$  simultaneous equations in  $M$  current endogenous variables and  $K$  exogenous variables, written as:

$$Y\Gamma + X\beta + U = 0, \quad (1.1)$$

where  $Y = [y_1 \cdots y_M]$  is  $NT \times M$ ,  $X = [x_1 \cdots x_K]$  is  $NT \times K$ , and  $U = [u_1 \cdots u_M]$  is  $NT \times M$ ,  $\Gamma = [\gamma_1^* \cdots \gamma_M^*]$  is  $M \times M$ , and  $\beta = [\beta_1^* \cdots \beta_K^*]$  is  $K \times M$ . Taking account of the normalization rule and the zero restrictions, a typical structural equation, say the  $j$ th one, can be written as:

$$y_j = Y_j\alpha_j + X_j\beta_j + u_j = Z_j\delta_j + u_j, \quad (1.2)$$

where  $Y_j = Y\tilde{H}_j$  is  $NT \times (\tilde{M}_j - 1)$ ,  $X_j = XL_j$  is  $NT \times K_j$ ,  $\alpha_j$  is  $(\tilde{M}_j - 1) \times 1$ ,  $\beta_j$  is  $K_j \times 1$ ,  $Z_j = [Y_j X_j]$ ,  $\delta_j = [\alpha_j' \beta_j']$ , with  $\tilde{H}_j$  and  $L_j$  being appropriate selection matrices.

The error components structure is given by:

$$u_j = (I_N \otimes \iota_T)\mu_j + (\iota_N \otimes I_T)\lambda_j + v_j \quad j = 1, \dots, M \quad (1.3)$$

where  $I_N$ ,  $I_T$  are identity matrices of order  $N$  and  $T$ , respectively,  $\iota_N$  and  $\iota_T$  are unit vectors of order  $N$  and  $T$ , respectively, and  $\otimes$  denotes the Kronecker product.  $\mu_j' = (\mu_{1j} \cdots \mu_{Nj})$ ,  $\lambda_j' = (\lambda_{1j} \cdots \lambda_{Tj})$ , and  $v_j' = (v_{11j} \cdots v_{NTj})$  are random vectors independent two by two with the following properties:

$$E(\mu_j) = 0, \quad E(\lambda_j) = 0, \quad E(v_j) = 0, \quad j = 1, \dots, M \quad (1.4)$$

$$E(\mu_j\mu_l') = \sigma_{\mu jl}^2 I_N, \quad E(\lambda_j\lambda_l') = \sigma_{\lambda jl}^2 I_T,$$

$$E(v_j v_l') = \sigma_{v jl}^2 I_{NT}, \quad \text{for } j \text{ and } l = 1, \dots, M. \quad (1.5)$$

Thus it follows that:

$$\Sigma = E(\text{vec } U)(\text{vec } U)' = \Sigma_\mu \otimes A + \Sigma_\lambda \otimes B + \Sigma_v \otimes I_{NT}, \quad (1.6)$$

where  $A = I_N \otimes \iota_T \iota_T'$ ,  $B = \iota_N \iota_N' \otimes I_T$ ,  $\Sigma_\mu = [\sigma_{\mu j l}^2]$ ,  $\Sigma_\lambda = [\sigma_{\lambda j l}^2]$ ,  $\Sigma_\nu = [\sigma_{\nu j l}^2]$ . We denote a typical block of  $\Sigma$ , corresponding to  $E(u_j u_i')$ , by  $\Sigma_{ji}$ .

The spectral decomposition of  $\Sigma$  is given by:

$$\Sigma = \sum_{i=1}^4 \Sigma_i \otimes M_i, \quad (1.7)$$

where  $\Sigma_1 = \Sigma_\nu + T \Sigma_\mu$ ,  $\Sigma_2 = \Sigma_\nu + N \Sigma_\lambda$ ,  $\Sigma_3 = \Sigma_\nu + T \Sigma_\mu + N \Sigma_\lambda$ ,  $\Sigma_4 = \Sigma_\nu$ ,  $M_1 = \frac{A}{T} - \frac{J_{NT}}{NT}$ ,  $M_2 = \frac{B}{N} - \frac{J_{NT}}{NT}$ ,  $M_3 = \frac{J_{NT}}{NT}$ ,  $M_4 = Q = I_{NT} - \frac{A}{T} - \frac{B}{N} + \frac{J_{NT}}{NT}$  with  $J_{NT} = \iota_{NT} \iota_{NT}'$ .  $M_1, M_2, M_3, M_4$  are idempotent matrices of rank  $m_1 = (N-1)$ ,  $m_2 = (T-1)$ ,  $m_3 = 1$ , and  $m_4 = (N-1)(T-1)$ , respectively, and  $M_1 + M_2 + M_3 + M_4 = I_{NT}$ . Given the particular form of equation (1.7), the inverse and determinant of  $\Sigma$  are (cf. Lemma 2.1, [12]):

$$\Sigma^{-1} = \sum_{i=1}^4 \Sigma_i^{-1} \otimes M_i, \quad |\Sigma| = \prod_{i=1}^4 |\Sigma_i|^{m_i}. \quad (1.8)$$

Now, the reduced form is obtained by solving equation (1.1) for  $Y$ :

$$Y = X\Pi + V, \quad (1.9)$$

where  $\Pi = -\beta\Gamma^{-1}$  and  $V = -U\Gamma^{-1}$ . Since  $\text{vec } V = -(\Gamma^{-1} \otimes I)' \text{vec } U$ , the variance-covariance structure of the reduced form is also of error components structure. In fact, we obtain:

$$\Omega = E(\text{vec } V)(\text{vec } V)' = \Omega_1 \otimes M_1 + \Omega_2 \otimes M_2 + \Omega_3 \otimes M_3 + \Omega_4 \otimes M_4, \quad (1.10)$$

where

$$\Omega_i = (\Gamma^{-1})' \Sigma_i \Gamma^{-1}. \quad (1.11)$$

Here, it may be added that the relationships that exist between  $\Sigma_1, \Sigma_2, \Sigma_3, \Sigma_4$ , and  $\Sigma_\mu, \Sigma_\lambda, \Sigma_\nu$  hold in the case of the reduced form also, giving  $\Omega_\mu, \Omega_\nu, \Omega_\lambda$  from  $\Omega_1, \Omega_2, \Omega_3, \Omega_4$  and vice versa.

The inverse and determinant of  $\Omega$  are analogous to those given for  $\Sigma$  in (1.8). It suffices to replace  $\Sigma_i$  by  $\Omega_i$ .

### Additional Assumptions

In deriving the limiting distribution of the various estimators, we further make the following assumptions:

- i. Independence between uncorrelated elements of each error component vector  $\mu_j, \lambda_j$  and  $\nu_j$ .

ii.  $X$  is a non-stochastic matrix such that:

$$\lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \frac{1}{NT} X' Q X = S, \quad (1.12)$$

$S$  being a positive definite matrix and  $Q$  being equal to  $M_4$  as defined previously. However, some precaution must be taken when the model contains a constant term. In this case, the matrix  $X' Q X$  is necessarily singular for any sample size and therefore its limit cannot be positive definite. Hence, we shall assume that either there is no constant term in the equations (and all the following results hold as such) or that the data are centered (and the results concern only the slope coefficients). It is worth noting that, with the covariance structure adopted in this paper, it is always possible to center the data before applying the GLS or the maximum likelihood principle, without modifying the results for the slope parameters.

iii.

$$\lim_{\substack{N \rightarrow \infty \\ T \rightarrow \infty}} \frac{N}{T} = 1. \quad (1.13)$$

For the maximum likelihood estimation, we obviously assume normality of the error components  $\mu_j$ ,  $\lambda_j$ , and  $\nu_j$  for  $j = 1, \dots, M$ .

## 2. MAXIMUM LIKELIHOOD ESTIMATION OF THE UNRESTRICTED REDUCED FORM

Before examining the maximum likelihood estimator, let us write down two obvious estimators of  $\Pi$ , namely, the feasible GLS estimator ( $\hat{\Pi}_{\text{GLS}}$ ) and the covariance estimator ( $\hat{\Pi}_{\text{cov}}$ ):

$$\text{vec}(\hat{\Pi}_{\text{GLS}}) = \left[ \sum_{i=1}^4 (\hat{\Omega}_i^{-1} \otimes X' M_i X) \right]^{-1} \sum_{i=1}^4 (\hat{\Omega}_i^{-1} \otimes X' M_i) \text{vec } Y, \quad (2.1)$$

$$\text{vec}(\hat{\Pi}_{\text{cov}}) = [I \otimes X' Q X]^{-1} (I \otimes X' Q) \text{vec } Y, \quad Q = M_4 \quad (2.2)$$

or

$$\hat{\Pi}_{\text{cov}} = (X' Q X)^{-1} X' Q Y. \quad (2.3)$$

The estimators of the variance components are obtained by ANOVA methods using residuals of the covariance method:

$$\hat{\Omega}_i = \frac{1}{m_i} (Y - X \hat{\Pi}_{\text{cov}})' M_i (Y - X \hat{\Pi}_{\text{cov}}) \quad i = 1, 2, \text{ and } 4, \quad (2.4)$$

$$\hat{\Omega}_3 = \hat{\Omega}_1 + \hat{\Omega}_2 - \hat{\Omega}_4. \quad (2.5)$$

There is no guaranty that the matrix in (2.5) is positive definite in a given sample. However, if we assume that the data are centered,  $X' M_3 X$  is equal to zero and the sum in (2.1) runs only over the indices 1, 2, and 4 and (2.5)

is not needed. Note also that in the case of only individual effects, there is no such problem.

It can be easily verified that both  $\sqrt{NT}(\text{vec } \hat{\Pi}_{\text{fGLS}} - \text{vec } \Pi)$  and  $\sqrt{NT}(\text{vec } \hat{\Pi}_{\text{cov}} - \text{vec } \Pi)$  have a normal limiting distribution with zero mean and covariance matrix equal to  $(\Omega_4 \otimes S^{-1})$ . This limiting distribution is the same as that of the full GLS estimator.

Let us add that our feasible GLS estimator is the same as the second one given in Baltagi [6], i.e., the one that uses covariance residuals for estimating the variance components. Avery [2] uses OLS residuals and this method also leads to an asymptotically efficient estimator of coefficients, as pointed out by Prucha [17].

We now proceed to develop the maximum likelihood estimator. The log likelihood, apart from an irrelevant constant, is

$$\log L(Y/\Pi, \Omega_\nu, \Omega_\mu, \Omega_\lambda) = -\frac{1}{2} \sum_{i=1}^4 m_i \log |\Omega_i| - \frac{1}{2} \sum_{i=1}^4 \text{tr } V' M_i V \Omega_i^{-1}. \quad (2.6)$$

Notice that we parametrize the likelihood function in terms of  $\Omega_\nu, \Omega_\mu, \Omega_\lambda$  (which are fixed parameters independent of  $T$  and  $N$ ), but we use the  $\Omega_i$  as short-hand notation. The definitions of  $\Omega_i$  are the same as the corresponding ones for  $\Sigma_i$  given immediately after formula (1.7). Also  $V$  stands for  $Y - X\Pi$ .

The log likelihood has to be maximized with respect to all the parameters, subject to the symmetry conditions

$$C \text{vec } \Omega_j = 0, \quad j = \mu, \lambda, \nu, \quad (2.7)$$

where  $C$  is a known  $\frac{M(M-1)}{2} \times M^2$  matrix of full row rank such that

$C'C$  is equal to the idempotent matrix  $(I - P)$ ,  $P$  being the commutation matrix. By writing the symmetry condition as in (2.7) above, we avoid extracting the redundant elements of the different covariance matrices.<sup>1</sup>

It turns out that the symmetry conditions are automatically satisfied by the first-order conditions of the associated Lagrangian function. We can therefore proceed to the direct maximization of  $\log L$ . It should be kept in mind, however, that the constraints represented by (2.7) are relevant for the computation of the (bordered) information matrix.

The first-order differential of (2.6) is<sup>2</sup>:

$$\begin{aligned} d \log L = & -\frac{1}{2} \sum_{i=1}^4 m_i \text{tr } \Omega_i^{-1} (d\Omega_i) \\ & - \sum_{i=1}^4 \frac{1}{2} \text{tr } V' M_i (dV) \Omega_i^{-1} \\ & + \frac{1}{2} \sum_{i=1}^4 \text{tr } V' M_i V \Omega_i^{-1} (d\Omega_i) \Omega_i^{-1}. \end{aligned} \quad (2.8)$$

Substituting

$$d\Omega_1 = d\Omega_\nu + Td\Omega_\mu, \quad (2.9)$$

$$d\Omega_2 = d\Omega_\nu + Nd\Omega_\lambda, \quad (2.10)$$

$$d\Omega_3 = d\Omega_\nu + Td\Omega_\mu + Nd\Omega_\lambda, \quad (2.11)$$

$$d\Omega_4 = d\Omega_\nu, \quad (2.12)$$

$$dV = Xd\Pi, \quad (2.13)$$

into the first-order condition  $d \log L = 0$  for all  $d\Pi \neq 0$  and  $d\Omega_j \neq 0$ ,  $j = \mu, \lambda, \nu$ , we can conveniently set up the following equations:

$$\text{vec } \Pi = \left[ \sum_{i=1}^4 (\Omega_i^{-1} \otimes X' M_i X) \right]^{-1} \sum_{i=1}^4 (\Omega_i^{-1} \otimes X' M_i) \text{vec } Y, \quad (2.14)$$

$$\Omega_i = \frac{1}{m_i} (Y - X\Pi)' M_i (Y - X\Pi) + \frac{1}{m_i} \Omega_i W \Omega_i, \quad i = 1, 2 \quad (2.15)$$

$$\Omega_4 = \frac{1}{m_4} (Y - X\Pi)' M_4 (Y - X\Pi) - \frac{1}{m_4} \Omega_4 W \Omega_4, \quad (2.16)$$

where

$$W = \Omega_3^{-1} (Y - X\Pi)' M_3 (Y - X\Pi) \Omega_3^{-1} - m_3 \Omega_3^{-1},$$

$$\Omega_3 = \Omega_1 + \Omega_2 - \Omega_4. \quad (2.17)$$

Contrary to the case of only individual effects ( $\Omega_\lambda = 0$ , see Magnus [12]), in the present situation no explicit expressions for the  $\Omega_i$  in terms of  $\Pi$  can be obtained. However, the whole system (2.14) to (2.17) can be solved iteratively as follows<sup>3</sup>:

STEP 1. Initial conditions:  $\Omega_i^{-1} = 0$ ,  $i = 1, 2, 3$  and  $\Omega_4 = I$ .

STEP 2. Use (2.14) to compute  $\text{vec } \Pi$ .

STEP 3. Compute  $\Omega_i$ ,  $i = 1, 2, 4$  from (2.15) and (2.16), using on the right-hand side of these equations the current estimates for  $\Pi$  and the previous ones for  $\Omega_i$ . (Note that in the first iteration  $W = 0$ ). Compute also  $\Omega_3 = \Omega_1 + \Omega_2 - \Omega_4$  and  $W$  from (2.17).

STEP 4. Go back to Step 2 until convergence is reached. If the above initial values are chosen, then it is easily verified that:

$\hat{\Pi} = \hat{\Pi}_{\text{cov}}$  at the first iteration, and

$\hat{\Pi} = \hat{\Pi}_{\text{fGLS}}$  at the second iteration.

### The Limiting Distribution

The limiting distribution of the maximum likelihood (ML) estimators can be said to be normal by virtue of Vickers' theorems (as stated in Magnus [13], p. 295). Indeed, it can be shown that:

- a. Each element of the information matrix, when divided by an appropriate normalizing factor, tends to a finite function of the true parameters as  $N$  and  $T$  tend to infinity; and
- b. The variance of each element of the matrix of second-order derivatives of the log-likelihood function, when divided by the square of the corresponding normalizing factor, converges uniformly to zero.

The moments of the limiting distribution can be obtained directly from the inverse of the bordered (limit) information matrix, see Appendix A. We may therefore conclude that:

$$\sqrt{NT} \text{vec}(\hat{\Pi}_{ML} - \Pi) \sim N(0, \Omega_\nu \otimes S^{-1}),$$

$$\sqrt{NT} \text{vec}(\hat{\Omega}_\nu - \Omega_\nu) \sim N\left(0, \frac{1}{2} (I + P)(2\Omega_\nu \otimes \Omega_\nu) \frac{1}{2} (I + P)\right),$$

$$\sqrt{N} \text{vec}(\hat{\Omega}_\mu - \Omega_\mu) \sim N\left(0, \frac{1}{2} (I + P)(2\Omega_\mu \otimes \Omega_\mu) \frac{1}{2} (I + P)\right),$$

$$\sqrt{T} \text{vec}(\hat{\Omega}_\lambda - \Omega_\lambda) \sim N\left(0, \frac{1}{2} (I + P)(2\Omega_\lambda \otimes \Omega_\lambda) \frac{1}{2} (I + P)\right).$$

The asymptotic equivalence between  $\text{vec } \hat{\Pi}_{ML}$ ,  $\text{vec } \hat{\Pi}_{cov}$ , and  $\text{vec } \hat{\Pi}_{fGLS}$  is thus proved.

### 3. "GENERALIZED" THREE-STAGE LEAST SQUARES

In this section we consider different structural estimation methods. Before discussing what we call "generalized" 3SLS (G3SLS) which is a full-information system method, a single equation method, namely, "generalized" 2SLS (G2SLS) is presented and these estimators are compared with the estimators obtained by indirect estimation when the structural equation in question is just-identified. Then the generalized 3SLS is presented as a direct extension of the generalized 2SLS and the asymptotic properties of the G3SLS are derived. Finally, it is shown that G3SLS reduces to G2SLS in the case of a just-identified system.

#### Generalized Two-Stage Least Squares

Let us consider the first structural equation and premultiply it by  $X'F^*$  where  $F^*$  is an  $NT \times NT$  matrix to be determined:

$$X'F^*y_1 = X'F^*Z_1\delta_1 + X'F^*u_1. \quad (3.1)$$

Let  $\hat{\delta}_{1(g)}$  denote the estimator of  $\delta_1$  obtained by applying GLS on (3.1). It can be shown that (see Balestra [3])  $F^* = \Sigma_{11}^{-1}$  minimizes the trace and determinant of the asymptotic covariance matrix of  $\hat{\delta}_{1(g)}$  and also gives the minimal positive definite asymptotic covariance matrix. This leads us to define the G2SLS estimator of  $\delta_1$  as follows:

$$\hat{\delta}_{1,G2SLS} = [Z_1 \Sigma_{11}^{-1} X(X' \Sigma_{11}^{-1} X)^{-1} X' \Sigma_{11}^{-1} Z_1]^{-1} \times Z_1' \Sigma_{11}^{-1} X(X' \Sigma_{11}^{-1} X)^{-1} X' \Sigma_{11}^{-1} y_1. \quad (3.2)$$

If we put  $F^* = Q$ , we get the 2SLS covariance estimator of  $\delta_1$ , namely:

$$\hat{\delta}_{1,cov} = [Z_1' QX(X' QX)^{-1} X' QZ_1]^{-1} Z_1' QX(X' QX)^{-1} X' Qy_1. \quad (3.3)$$

Another 2SLS estimator is given by

$$\hat{\delta}_{1,RF} = (\hat{Z}_1' Q \hat{Z}_1)^{-1} \hat{Z}_1' Qy_1, \quad (3.4)$$

where  $Z_1 = [\hat{Y}_1 X_1]$ ,  $\hat{Y}_1 = X \hat{\Pi} \tilde{H}_1$ ,  $\hat{\Pi}$  being any consistent estimator of the reduced form. Straight calculation shows that when  $\Pi$  is estimated by the covariance method, the two estimators in (3.3) and (3.4) are equal.

The G2SLS is not feasible, in general, since the variances are not known. These variances can be consistently estimated in the following way. Use either  $\hat{\delta}_{1,cov}$  or  $\hat{\delta}_{1,RF}$  to compute residuals for the first structural equation, i.e.,  $\hat{u}_1 = y_1 - Z_1 \hat{\delta}_1$ . Then, by the ANOVA formulas, we get

$$\hat{\sigma}_{i11}^2 = \frac{1}{m_i} \hat{u}_1' M_i \hat{u}_1, \quad i = 1, 2, 4 \quad (3.5)$$

$$\hat{\sigma}_{311}^2 = \hat{\sigma}_{111}^2 + \hat{\sigma}_{211}^2 - \hat{\sigma}_{411}^2. \quad (3.6)$$

With these estimates we can compute  $\hat{\Sigma}_{11} = \sum_i \hat{\sigma}_{i11}^2 M_i$  and construct the feasible G2SLS estimator, i.e.,

$$\hat{\delta}_{1,fG2SLS} = [Z_1' \hat{\Sigma}_{11}^{-1} X(X' \hat{\Sigma}_{11}^{-1} X)^{-1} X' \hat{\Sigma}_{11}^{-1} Z_1]^{-1} Z_1' \hat{\Sigma}_{11}^{-1} X(X' \hat{\Sigma}_{11}^{-1} X)^{-1} X' \hat{\Sigma}_{11}^{-1} y_1. \quad (3.7)$$

It is interesting to note that the four 2SLS estimators presented here all share the same asymptotic distribution which is normal with zero mean and covariance matrix given by

$$\sigma_{411}^2 [(\Pi \tilde{H}_1 L_1)' S (\Pi \tilde{H}_1 L_1)]^{-1}. \quad (3.8)$$

Their small sample properties are not known yet; they could be investigated by Monte Carlo experiments.

In addition, when the equation is just-identified, the ILS estimator of  $\delta_1$  derived using  $\hat{\Pi}_{cov}$  is equal to the 2SLS covariance estimator and the one derived using either  $\hat{\Pi}_{cov}$  or  $\hat{\Pi}_{fGLS}$  has the same limiting distribution as that of the feasible G2SLS estimator.

### Generalized Three-Stage Least Squares

The extension from G2SLS to G3SLS is made in an analogous manner to that from classical 2SLS to 3SLS. Let

$$Z_* = \begin{bmatrix} Z_1 & & 0 \\ & \ddots & \\ 0 & & Z_M \end{bmatrix}, \quad \delta = \begin{bmatrix} \delta_1 \\ \vdots \\ \delta_M \end{bmatrix}, \quad y = \text{vec } Y, \quad u = \text{vec } U.$$



Then, the set of  $M$  structural equations can be written compactly as:

$$y = Z_*\delta + u. \quad (3.9)$$

Let  $D = \text{diag}(\Sigma_{11} \cdots \Sigma_{MM})$  and let us premultiply (3.9) by  $\chi'D^{-1}$ , where  $\chi = I_M \otimes X$ . This gives:

$$\chi'D^{-1}y = \chi'D^{-1}Z_*\delta + \chi'D^{-1}u. \quad (3.10)$$

Performing GLS on (3.10) yields the G3SLS estimator of  $\delta$ :

$$\begin{aligned} \hat{\delta}_{\text{G3SLS}} &= (Z_*'D^{-1}\chi(\chi'D^{-1}\Sigma D^{-1}\chi)^{-1}\chi'D^{-1}Z_*)^{-1} \\ &\quad \times Z_*'D^{-1}\chi(\chi'D^{-1}\Sigma D^{-1}\chi)^{-1}\chi'D^{-1}y. \end{aligned} \quad (3.11)$$

Again, the unknown variances have to be estimated and one way of estimating them is as follows:

$$\hat{\sigma}_{ijl}^2 = \frac{1}{m_i} (y_j - Z_j\hat{\delta}_{j,\text{fG2SLS}})'M_i(y_l - Z_l\hat{\delta}_{l,\text{fG2SLS}}) \quad i = 1, 2, 4$$

and

$$\hat{\sigma}_{3jl}^2 = \hat{\sigma}_{1jl}^2 + \hat{\sigma}_{2jl}^2 - \hat{\sigma}_{4jl}^2,$$

whence

$$\hat{\Sigma}_{jl} = \sum_i \hat{\sigma}_{ijl}^2 M_i, \quad \hat{\Sigma} = [\hat{\Sigma}_{jl}], \quad \hat{D} = \text{diag}[\hat{\Sigma}_{11}, \dots, \hat{\Sigma}_{MM}],$$

and a feasible G3SLS estimator ( $\hat{\delta}_{\text{fG3SLS}}$ ) is obtained by substituting  $\hat{D}$ ,  $\hat{\Sigma}$  for  $D$ ,  $\Sigma$ , respectively, in (3.11).

It can be shown that  $\sqrt{NT}(\hat{\delta}_{\text{fG3SLS}} - \delta)$  has a normal limiting distribution with zero mean and  $(\bar{P}'(\Sigma_4^{-1} \otimes S)\bar{P})^{-1}$  as covariance matrix where

$$\bar{P} = \begin{bmatrix} (\Pi\tilde{H}_1L_1) & & 0 \\ & \ddots & \\ 0 & & (\Pi\tilde{H}_ML_M) \end{bmatrix}, \quad (3.12)$$

When the whole system is just-identified, the feasible G3SLS estimator of the coefficients of any structural equation is exactly equal to the corresponding feasible G2SLS estimator if we estimate its covariance matrix by the same method in both cases and their common limiting distribution is the same as that of the ILS estimator.

#### 4. FULL INFORMATION MAXIMUM LIKELIHOOD ESTIMATION OF THE STRUCTURAL FORM

In this section we shall discuss the ML estimation of the constrained structural form. Again it is extremely difficult to get analytical results since the

normal equations can be solved only by numerical methods. To this end, we adapt a procedure suggested by Pollock [16] for the classical simultaneous equation model. It is also shown that the FIML estimator has the same asymptotic distribution as the G3SLS estimator.

### The Problem

The starting point is the log-likelihood function, see equation (2.6), which is to be expressed now in terms of the structural parameters. We use the following identities and definitions:

$$\Omega_i = (\Gamma^{-1})' \Sigma_i \Gamma^{-1}$$

$$\Pi = -\beta \Gamma^{-1}$$

$$\Theta' = [\Gamma' \beta']$$

$$Z = [YX]$$

$$L' = [I0], \quad L'\Theta = \Gamma$$

where the  $\Sigma_i$  are defined in (1.7).

Then we can write

$$\begin{aligned} \log L(Y/\Theta, \Sigma_\nu, \Sigma_\mu, \Sigma_\lambda) = & -\frac{1}{2} \sum_{i=1}^4 m_i \log |\Sigma_i| + \frac{1}{2} NT \log |L'\Theta|^2 \\ & - \frac{1}{2} \text{tr} \sum_{i=1}^4 \Theta' Z' M_i Z \Theta \Sigma_i^{-1}. \end{aligned} \quad (4.1)$$

We are interested in maximizing this function with respect to  $\Theta, \Sigma_\nu, \Sigma_\mu$ , and  $\Sigma_\lambda$  subject to two types of restrictions:

- i. A priori restrictions on the coefficients. These must include the normalization rule for each equation (typically  $\gamma_{ii} = -1$ ), the exclusion restrictions (zero a priori coefficients) and eventually other linear restrictions on the parameters. They can be expressed conveniently by

$$R \text{vec } \Theta = r,$$

where  $R$  is a known  $s \times M(M+K)$  matrix of full row rank and  $r$  is a known  $s \times 1$  vector of constants. To ensure identification,  $s$  must be greater than or equal to  $M^2$ .

- ii. Symmetry conditions:

$$C \text{vec } \Sigma_j = 0 \quad j = \nu, \mu, \lambda$$

where  $C$  is the matrix defined in (2.7).

Again, it can be verified that the first-order conditions for maximization automatically satisfy the symmetry conditions. These can therefore be neglected for this purpose, but they will be important for the computation of the information matrix.

We therefore write the following Lagrangian function:

$$L^* = \log L + \Lambda' (r - R \text{vec } \Theta), \quad (4.2)$$

where  $\Lambda$  is a vector of Lagrangian multipliers.

### The ML Solution

The first-order differential of the Lagrangian function is given by:

$$\begin{aligned} dL^* = & -\frac{1}{2} \sum_i m_i \text{tr } \Sigma_i^{-1} d\Sigma_i + NT \text{tr}(\Theta' L)^{-1} L' d\Theta \\ & - \frac{1}{2} \text{tr} \sum_i (d\Theta)' Z' M_i Z \Theta \Sigma_i^{-1} - \frac{1}{2} \text{tr} \sum_i \Theta' Z' M_i Z d\Theta \Sigma_i^{-1} \\ & + \frac{1}{2} \text{tr} \sum_i \Theta' Z' M_i Z \Theta \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} \\ & + \Lambda' R d \text{vec } \Theta + (d\Lambda)' R \text{vec } \Theta. \end{aligned} \quad (4.3)$$

Once again, substituting

$$d\Sigma_1 = d\Sigma_\nu + T d\Sigma_\mu, \quad (4.4)$$

$$d\Sigma_2 = d\Sigma_\nu + N d\Sigma_\lambda, \quad (4.5)$$

$$d\Sigma_3 = d\Sigma_\nu + T d\Sigma_\mu + N d\Sigma_\lambda, \quad (4.6)$$

$$d\Sigma_4 = d\Sigma_\nu, \quad (4.7)$$

$$(\Theta' L)^{-1} = \Gamma'^{-1} = \Omega_4 \Gamma \Sigma_4^{-1} = \Omega_4 L' \Theta \Sigma_4^{-1}, \quad (4.8)$$

into the first-order condition which is  $dL^* = 0$  for all  $d \text{vec } \Theta \neq 0$ ,  $d\Lambda \neq 0$ , and  $d \text{vec } \Sigma_j \neq 0$ ,  $j = \mu, \nu, \lambda$  and using the familiar vec-trace relationship, we can derive the following system of normal equations:

$$W \text{vec } \Theta - R' \Lambda = 0, \quad (4.9)$$

$$R \text{vec } \Theta = r, \quad (4.10)$$

$$\Sigma_4 = \frac{1}{m_4} \Theta' Z' M_4 Z \Theta - \frac{1}{m_4} \Sigma_4 \tilde{W} \Sigma_4, \quad (4.11)$$

$$\Sigma_1 = \frac{1}{m_1} \Theta' Z' M_1 Z \Theta + \frac{1}{m_1} \Sigma_1 \tilde{W} \Sigma_1, \quad (4.12)$$

$$\Sigma_2 = \frac{1}{m_2} \Theta' Z' M_2 Z \Theta + \frac{1}{m_2} \Sigma_2 \tilde{W} \Sigma_2, \quad (4.13)$$

where

$$W = \Sigma_4^{-1} \otimes NTL\Omega_4 L' - \sum_{i=1}^4 (\Sigma_i^{-1} \otimes Z' M_i Z),$$

$$\tilde{W} = \Sigma_3^{-1} \Theta' Z' M_3 Z \Theta \Sigma_3^{-1} - m_3 \Sigma_3^{-1}.$$

The maximum likelihood estimates are obtained by solving simultaneously equations (4.9) to (4.13) together with the two definitions:

$$\Sigma_3 = \Sigma_1 + \Sigma_2 - \Sigma_4, \quad (4.14)$$

$$\Omega_4 = (\Gamma')^{-1} \Sigma_4 \Gamma^{-1}. \quad (4.15)$$

This system of equations is highly nonlinear.<sup>4</sup> We notice, however, that an explicit solution can be found for  $\text{vec } \Theta$  in terms of the different covariance matrices. Equations (4.9) and (4.10) can be combined to yield:

$$\begin{bmatrix} W & R' \\ R & 0 \end{bmatrix} \begin{bmatrix} \text{vec } \Theta \\ -\Lambda \end{bmatrix} = \begin{bmatrix} 0 \\ r \end{bmatrix}. \quad (4.16)$$

The first matrix on the left-hand side of (4.16) is nonsingular, if and only if

- i. Rank  $(R') = s$ , which is true by hypothesis, and
- ii. Rank  $(I - R'(RR')^{-1}R)W(I - R'(RR')^{-1}R) = M(M + K) - s$ , which is satisfied whenever the conditions for identification are met. Its inverse (see Balestra [4]) is given by

$$\begin{bmatrix} H_1 & H_2 \\ H_2' & H_3 \end{bmatrix},$$

where

$$H_1 = F(F'WF)^{-1}F',$$

$$H_2 = -F(F'WF)^{-1}F'WR'(RR')^{-1} + R'(RR')^{-1},$$

$$H_3 = -(RR')^{-1}RWR'(RR')^{-1} + (RR')^{-1}RWF(F'WF)^{-1}F'WR'(RR')^{-1}$$

and  $F$  is an  $M(K + M) \times (M^2 + MK - s)$  matrix of orthonormal vectors such that  $FF' = I - R'(RR')^{-1}R'$ . We therefore obtain the following solution:

$$\text{vec } \Theta = H_2 r = -F(F'WF)^{-1}F'WR'(RR')^{-1}r + R'(RR')^{-1}r. \quad (4.17)$$

It is very useful to note, from an operational point of view, that whenever the usual restrictions are considered only (normalization and exclusion), the matrix  $R$  can be partitioned as

$$R = \begin{bmatrix} R_1 & & 0 \\ & R_2 & \\ & & \ddots \\ 0 & & & R_M \end{bmatrix}, \quad (4.18)$$

when  $R_i$  is an  $s_i \times (M + K)$  matrix whose rows are elementary vectors. Therefore,  $RR' = I_s$ . Likewise, the corresponding part of  $r$ ,  $r_i$ , is an elementary vector (with a minus sign in front) and  $R_i' r_i = r_i$ . Furthermore, the matrix  $F'$  can also be partitioned in a block diagonal form

$$F' = \begin{bmatrix} F'_1 & & 0 \\ & F'_2 & \\ & & \ddots \\ 0 & & & F'_M \end{bmatrix}, \quad (4.19)$$

where the  $M + K - s_i$  rows of  $F_i$  are just the elementary vectors which are complementary (orthogonal) to those appearing in  $R_i$ . The matrix  $F$  is therefore computed without any difficulty. In this case, the expression for  $\text{vec } \Theta$  simplifies to:

$$\text{vec } \Theta = -F(F'WF)^{-1}F'Wr + r, \quad (4.20)$$

and the non-constrained coefficients are obtained by premultiplication by  $F'$ , ( $F'F = I$ ,  $F'r = 0$ ), i.e.,

$$F' \text{vec } \Theta = -(F'WF)^{-1}F'Wr. \quad (4.21)$$

We therefore suggest the following iterative procedure for the solution of the normal equations:

STEP 1. Initial conditions:

$$\Sigma_i^{-1} = 0, \quad i = 1, 2, 3 \text{ (their limits)}$$

$$\Sigma_4 = I$$

$$\Omega_4 = \frac{1}{NT} (Y - X\hat{\Pi})'Q(Y - X\hat{\Pi})$$

where  $\hat{\Pi}$  is a consistent estimator of the reduced form, say  $\hat{\Pi}_{\text{cov}}$ .

STEP 2. Use (4.17) or, in case of usual restrictions, use (4.20) to estimate  $\text{vec } \Theta$ .

STEP 3. Compute  $\Sigma_i$ ,  $i = 1, 2, 4$ , from (4.11), (4.12), and (4.13) using on the right-hand side the current estimate for  $\Theta$  and the old ones (from the pre-

vious iteration) for the different  $\Sigma_i$ . Compute  $\Sigma_3$  from (4.14) and  $\Omega_4$  from (4.15).

STEP4. Go back to Step 2 until convergence is reached.

In the case of usual restrictions and with the initial conditions stated in Step 1, the matrix  $W$  becomes

$$W = I \otimes -G,$$

where

$$G = Z'QX(X'QX)^{-1}X'QZ.$$

In view of the block diagonal form of  $F'$  we can obtain directly, for Step 2:

$$\text{vec } \Theta = -F[\text{diag}(F_i'GF_i)^{-1}][\text{diag}(F_i'G)]r + r$$

which, for the coefficients of the  $i$ th equation, gives

$$\Theta_i = -F_i(F_i'GF_i)^{-1}F_i'Gr_i + r_i. \quad (4.22)$$

Now  $F_i'Z = Z_i$  where  $Z_i$  contains all the explanatory variables of the  $i$ th equation (both endogenous and exogenous) and  $Zr_i = -y_i$ , the explained variable. Therefore, we get:

$$\Theta_i = F_i[Z_i'QX(X'QX)^{-1}X'QZ_i]^{-1}Z_i'QX(X'QX)^{-1}X'Qy_i + r_i, \quad (4.23)$$

which is seen to be identical (for the non-constrained coefficients) to the 2SLS covariance estimator, see (3.3). Hence, the first iteration gives the 2SLS solution. At the second iteration, if we keep the initial values for  $\Sigma_1^{-1}$ ,  $\Sigma_2^{-1}$ ,  $\Sigma_3^{-1}$ , and  $\Omega_4$ , and compute  $\Sigma_4$  according to (4.11), then we obtain a 3SLS estimator of the covariance type (which is a particular case of our G3SLS with  $D^{-1}$  replaced by  $I_M \otimes Q$ ).

### The Limiting Distribution

As in the case of the reduced form, the requirements for the application of Vickers' theorems are fulfilled and, therefore, the FIML estimators are asymptotically normally distributed. The moments of the limiting distribution of these estimators are easily obtained from the inverse of the limit of the bordered information matrix. See Appendix B for the computation of the information matrix and its inverse. We obtain that  $\sqrt{NT} \text{vec}(\hat{\Theta}_{ML} - \Theta)$  has a limiting normal distribution with zero mean and covariance matrix equal to

$$F[F'(\Sigma_v^{-1} \otimes D)F]^{-1}F', \quad (4.24)$$

where  $F$  is a matrix of orthonormal vectors such that  $FF' = I - R'(RR')^{-1}R$  and  $D$  is given by

$$D = \begin{bmatrix} \Pi' \\ I \end{bmatrix} S[\Pi I], \quad S = \lim_{NT} \frac{1}{NT} X'QX. \quad (4.25)$$

We can write (4.24) in the following form:

$$F \left[ F' \left( I \otimes \begin{bmatrix} \Pi' \\ I \end{bmatrix} \right) (\Sigma_\nu^{-1} \otimes S) (I \otimes [\Pi]) F \right]^{-1} F'. \quad (4.26)$$

Now, when the a priori restrictions are just the zero restrictions and the normalization rule, then

$$F' \left( I \otimes \begin{bmatrix} \Pi' \\ I \end{bmatrix} \right) = \tilde{P}',$$

where  $\tilde{P}$  is defined in (3.12) and the formula in (4.26), for the unconstrained coefficients, is equal to the covariance matrix of the limiting distribution of the G3SLS estimator. It follows that the FIML estimator and the G3SLS estimator are asymptotically equivalent.

For the ML estimators of the variance components, we have the following limiting distributions:

$$\sqrt{NT} \text{vec}(\hat{\Sigma}_\nu - \Sigma_\nu) \sim N(0, H^{22}),$$

$$\sqrt{N} \text{vec}(\hat{\Sigma}_\mu - \Sigma_\mu) \sim N(0, \frac{1}{2} (I + P)(2\Sigma_\mu \otimes \Sigma_\mu) \frac{1}{2} (I + P)),$$

$$\sqrt{T} \text{vec}(\hat{\Sigma}_\lambda - \Sigma_\lambda) \sim N(0, \frac{1}{2} (I + P)(2\Sigma_\lambda \otimes \Sigma_\lambda) \frac{1}{2} (I + P)),$$

where

$$\begin{aligned} H^{22} = & \frac{1}{2} (I + P)(2\Sigma_\nu \otimes \Sigma_\nu) \frac{1}{2} (I + P) \\ & + \frac{1}{2} (I + P)(2I \otimes \Sigma_\nu (L'\Theta)^{-1} L') F [F' (\Sigma_\nu^{-1} \otimes D) F]^{-1} \\ & \times F' (2I \otimes L(\Theta' L)^{-1} \Sigma_\nu) \frac{1}{2} (I + P). \end{aligned}$$

## 5. CONCLUSION

The maximum likelihood approach developed in this paper for the case of a system of simultaneous equations with error component structure is both conceptually simple and operationally interesting. The elegant solution proposed by Pollock for the classical case still holds in the present context with only a moderate increase in the complexity of the analytical expressions and in the computational burden.

In addition, the approach permits a unified study of the different full information estimators available and establishes the asymptotic equivalence between the ML estimator and the G3SLS estimator. Finally, although no explicit proof is given in this paper, it can be shown in the present context

also that the limited information maximum likelihood (LIML) estimator is equal to the FIML estimator of a reduced system consisting of the structural equation in question and the reduced form equations for its explanatory endogenous variables.

## NOTES

1. This is an alternative approach to the one based on the elimination matrix as adopted, for instance, by Magnus and Neudecker [14] and Balestra [3] or on the equivalent vech operator (which stacks the columns of a symmetric matrix starting each column at its diagonal element) as in Henderson and Searle [10]. Our procedure is inspired by an article by F.J.H. Don [8].

2. See Magnus [12] for the use of the differential technique.

3. In the case of only individual effects, we will have a zigzag iterative procedure whose convergence to a solution of the ML first-order conditions may be proved using similar assumptions as in [15] and [9]. However, in the presence of both effects, the iterative procedure suggested here, as well as the one proposed later for the FIML estimation of the structural form, are of a more complex nature, as direct maximization of the log-likelihood function with respect to the covariance parameters given the coefficient parameters is not possible. In this case, the problem of convergence needs further investigation and the authors are currently examining it in greater detail.

4. For the case of no time effects, the log-likelihood function in (4.1) must be replaced by:

$$\log L = -\frac{1}{2} \sum_{1,4} m_i \log |\Sigma_i| + \frac{1}{2} NT \log |L' \Theta|^2 - \frac{1}{2} \text{tr} \sum_{1,4} \Theta' Z' M_i Z \Theta \Sigma_i^{-1},$$

with  $m_1 = N$ ,  $m_2 = N(T-1)$ ,  $M_1 = A/T$ , and  $M_4 = I - M_1$ . In this case the solution of the normal equations is much simpler, since an explicit solution can be found for  $\Sigma_1$  and  $\Sigma_4$  in terms of  $\Theta$ . Equations (4.9) and (4.10) remain the same, but with  $W$  defined as

$$W = \Sigma_4^{-1} \otimes NTL\Omega_4 L' - \sum_{1,4} (\Sigma_i^{-1} \otimes Z' M_i Z),$$

on the other hand, for the covariance matrices we have (instead of (4.11)–(4.13)):

$$\Sigma_4 = \frac{1}{m_4} \Theta' Z' M_4 Z \Theta \quad \text{and} \quad \Sigma_1 = \frac{1}{m_1} \Theta' Z' M_1 Z \Theta,$$

together with the definition  $\Omega_4 = (\Gamma')^{-1} \Sigma_4 \Gamma^{-1}$ .

## REFERENCES

1. Amemiya, T. The estimation of variances in a variance-components model. *International Economic Review* 12 (1971): 1–13.
2. Avery, R.B. Error components and seemingly unrelated regressions. *Econometrica* 45 (1977): 199–209.
3. Balestra, P. La dérivation matricielle. *Collection de l'Institut de mathématiques économiques de Dijon* 12. Sirey, Paris, 1983.
4. Balestra, P. Determinant and inverse of a sum of matrices with applications in economics and statistics. *Document de travail* 24. Institut de mathématiques économiques de Dijon, France, 1978.
5. Balestra, P. & M. Nerlove. Pooling cross-section and time-series data in the estimation of a dynamic model: The demand for natural gas. *Econometrica* 34 (1966): 585–612.
6. Baltagi, B.H. On seemingly unrelated regressions with error components. *Econometrica* 48 (1980): 1547–1551.
7. Baltagi, B.H. Simultaneous equations with error components. *Journal of Econometrics* 17 (1981): 189–200.



8. Don, F.J.H. The use of generalized inverses in restricted maximum likelihood. *Linear Algebra and its Applications*, Vol. 70 (1985).
9. Don, F.J.H. & J.R. Magnus. On the unbiasedness of iterated GLS estimators. *Communications in Statistics* A9(5) (1980): 519–527.
10. Henderson, H.V. & S.R. Searle. Vec and vech operators for matrices, with some uses in Jacobians and multivariate statistics. *The Canadian Journal of Statistics* 7(1) (1979): 65–81.
11. Hendry, D.F. The structure of simultaneous equation estimators. *Journal of Econometrics* 4 (1976): 51–88.
12. Magnus, J.R. Multivariate error components analysis of linear and nonlinear regression models by maximum likelihood. *Journal of Econometrics* 19 (1982): 239–285.
13. Magnus, J.R. Maximum likelihood estimation of the GLS model with unknown parameters in the disturbance covariance matrix. *Journal of Econometrics* 7 (1978): 281–312.
14. Magnus, J.R. & H. Neudecker. The elimination matrix: Some theorems and applications. *SIAM Journal on Algebraic and Discrete Methods* 1 (1980): 422–449.
15. Oberhofer, W. & J. Kmenta. A general procedure for obtaining maximum likelihood estimates in generalized regression models. *Econometrica* 42 (1974): 579–590.
16. Pollock, D.S.G. *The algebra of econometrics*. New York: Wiley, 1979.
17. Prucha, I.R. On the asymptotic efficiency of feasible Aitken estimator for seemingly unrelated regression models with error components. *Econometrica* 52 (1984): 203–207.
18. Prucha, I.R. Maximum likelihood and instrumental variable estimation in simultaneous equation systems with error components. *International Economic Review*, 26(2) (1985): 491–506.
19. Varadharajan, J. Estimation of simultaneous linear equation models with error component structure. *Cahiers du Département d'économétrie* 81.06. Université de Genève, Switzerland, 1981.
20. Wallace, T.D. & A. Hussain. The use of error components models in combining cross-section with time-series data. *Econometrica* 37 (1969): 55–72.

## APPENDIX A: THE BORDERED INFORMATION MATRIX OF THE REDUCED FORM AND ITS INVERSE

The second-order differential of log  $L$  is:

$$\begin{aligned}
 d^2 \log L = & \frac{1}{2} \sum_{i=1}^4 m_i \operatorname{tr} \Omega_i^{-1} (d\Omega_i) \Omega_i^{-1} (d\Omega_i) \\
 & + \sum_{i=1}^4 \operatorname{tr} V' M_i (dV) \Omega_i^{-1} d\Omega_i \Omega_i^{-1} \\
 & - \sum_{i=1}^4 \operatorname{tr} (dV)' M_i (dV) \Omega_i^{-1} \\
 & - \sum_{i=1}^4 \operatorname{tr} V' M_i V \Omega_i^{-1} (d\Omega_i) \Omega_i^{-1} (d\Omega_i) \Omega_i^{-1} \\
 & + \sum_{i=1}^4 \operatorname{tr} V' M_i (dV) \Omega_i^{-1} (d\Omega_i) \Omega_i^{-1}.
 \end{aligned} \tag{A.1}$$

The negative of its expectation, noting that  $E(V'M_i dV) = 0$  and  $E(V'M_i V) = m_i \Omega_i$ , is:

$$E(-d^2 \log L) = \frac{1}{2} \sum_{i=1}^4 m_i \text{tr} \Omega_i^{-1} (d\Omega_i) \Omega_i^{-1} (d\Omega_i) + \sum_{i=1}^4 \text{tr}(d\Pi)' X' M_i X (d\Pi) \Omega_i^{-1}. \quad (\text{A.2})$$

Using the familiar vec-trace relationship in conjunction with the formulas (2.9) to (2.12) and ordering the parameters as  $[(\text{vec } \Pi)', (\text{vec } \Omega_\nu)', (\text{vec } \Omega_\mu)', (\text{vec } \Omega_\lambda)']$ , the information matrix  $\Psi$  is easily obtained as the symmetric matrix  $A$ , given on p. 241.

As noted by Amemiya [1] for the single equation model, we cannot simply take the limit of  $\Psi$  when divided by  $NT$ , but instead we must take the limit of  $\eta \Psi \eta$  where

$$\eta = \text{diag} \left[ \frac{1}{\sqrt{NT}} I, \frac{1}{\sqrt{NT}} I, \frac{1}{\sqrt{N}} I, \frac{1}{\sqrt{T}} I \right], \quad \text{partitioned appropriately.}$$

Noting the following limits (as both  $N$  and  $T$  tend to infinity)

$$\begin{aligned} \frac{1}{NT} X' M_i X &\rightarrow S, \\ \Omega_i^{-1} &\rightarrow 0, \quad i = 1, 2, 3 \\ T \Omega_1^{-1} &\rightarrow \Omega_\nu^{-1}, \\ N \Omega_2^{-1} &\rightarrow \Omega_\mu^{-1}, \\ T \Omega_3^{-1} &\rightarrow (\Omega_\mu + \Omega_\lambda)^{-1}, \quad \left( \frac{N}{T} \rightarrow 1 \right) \end{aligned}$$

we obtain the bordered information matrix:

$$H = \begin{bmatrix} \Omega_\nu^{-1} \otimes S & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{2} (\Omega_\nu^{-1} \otimes \Omega_\nu^{-1}) & 0 & 0 & C' & 0 & 0 \\ 0 & 0 & \frac{1}{2} (\Omega_\mu^{-1} \otimes \Omega_\mu^{-1}) & 0 & 0 & C' & 0 \\ 0 & 0 & 0 & \frac{1}{2} (\Omega_\lambda^{-1} \otimes \Omega_\lambda^{-1}) & 0 & 0 & C' \\ 0 & C & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & C & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & C & 0 & 0 & 0 \end{bmatrix}.$$

Due to its particularly simple form, the first block on the main diagonal of its inverse is obviously given by the inverse of the corresponding block. For the next three blocks on the main diagonal of the inverse of the information matrix, each block is the first block of the inverse of

Matrix A:

$$\begin{bmatrix} \sum_{i=1}^4 (\Omega_i^{-1} \otimes X' M_i X) & 0 & 0 & 0 \\ 0 & \frac{1}{2} \sum_{i=1}^4 m_i (\Omega_i^{-1} \otimes \Omega_i^{-1}) & & \\ 0 & \frac{T}{2} \sum_{i,3} m_i (\Omega_i^{-1} \otimes \Omega_i^{-1}) & \frac{T^2}{2} \sum_{i,3} m_i (\Omega_i^{-1} \otimes \Omega_i^{-1}) & \\ 0 & \frac{N}{2} \sum_{i,3} m_i (\Omega_i^{-1} \otimes \Omega_i^{-1}) & \frac{NT}{2} m_3 (\Omega_3^{-1} \otimes \Omega_3^{-1}) & \frac{N^2}{2} \sum_{i,3} m_i (\Omega_i^{-1} \otimes \Omega_i^{-1}) \end{bmatrix}$$

(symmetric)

Matrix B:

$$\begin{bmatrix} \left\{ NT[(L'\Theta)L' \otimes L'(\Theta'L)^{-1}]P - \sum_{i=1}^4 \Sigma_i^{-1} \otimes (D_i + m_i L \Omega_i L') \right\} & & & \\ \sum_{i=1}^4 m_i \Sigma_i^{-1} \otimes L(\Theta'L)^{-1} & \frac{1}{2} \sum_{i=1}^4 m_i (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) & & \\ \sum_{i,3} m_i T \Sigma_i^{-1} \otimes L(\Theta'L)^{-1} & \frac{T}{2} \sum_{i,3} m_i (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) & \frac{T^2}{2} \sum_{i,3} m_i (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) & \\ \sum_{i,3} m_i N \Sigma_i^{-1} \otimes L(\Theta'L)^{-1} & \frac{N}{2} \sum_{i,3} m_i (\Sigma_i^{-1} \otimes \Sigma_i^{-1}) & NT m_3 \Sigma_3^{-1} \otimes \Sigma_3^{-1} & \frac{N^2}{2} \sum_{i,3} m_i \Sigma_i^{-1} \otimes \Sigma_i^{-1} \end{bmatrix}$$

(symmetric)

$$\begin{bmatrix} \frac{1}{2}(\Omega_j^{-1} \otimes \Omega_j^{-1}) & C' \\ C & 0 \end{bmatrix}.$$

According to the usual inversion rule (see for instance Balestra [4, p. 10]), this can be expressed as

$$\bar{F} \left[ \bar{F}' \frac{1}{2} (\Omega_j^{-1} \otimes \Omega_j^{-1}) \bar{F} \right]^{-1} \bar{F}, \quad (\text{A.3})$$

for  $\bar{F}$  such that  $\bar{F}\bar{F}' = I - C'C = I - \frac{1}{2}(I - P)\frac{1}{2}(I + P)$ , where  $P$  is the commutation matrix. Note that  $\frac{1}{2}(I + P)$  is idempotent.

From the properties of the commutation matrix it is easily established that

$$\bar{F}\bar{F}'(2\Omega_j \otimes \Omega_j)\bar{F}\bar{F}' = \bar{F}\bar{F}'(2\Omega_j \otimes \Omega_j). \quad (\text{A.4})$$

Therefore, we obtain successively:

$$\bar{F}\bar{F}'(2\Omega_j \otimes \Omega_j)\bar{F}\bar{F}' \left( \frac{1}{2} \Omega_j^{-1} \otimes \Omega_j^{-1} \right) = \bar{F}\bar{F}',$$

$$\bar{F}\bar{F}'(2\Omega_j \otimes \Omega_j)\bar{F} \left[ \bar{F}' \left( \frac{1}{2} \Omega_j^{-1} \otimes \Omega_j^{-1} \right) \bar{F} \right] = \bar{F}, \quad (\bar{F}'\bar{F} = I)$$

$$\bar{F}\bar{F}'(2\Omega_j \otimes \Omega_j)\bar{F} = \bar{F} \left[ \bar{F}' \left( \frac{1}{2} \Omega_j^{-1} \otimes \Omega_j^{-1} \right) \bar{F} \right]^{-1},$$

$$\frac{1}{2}(I + P)(2\Omega_j \otimes \Omega_j)\frac{1}{2}(I + P) = \bar{F} \left[ \bar{F}' \left( \frac{1}{2} \Omega_j^{-1} \otimes \Omega_j^{-1} \right) \bar{F} \right]^{-1} \bar{F}.$$

The left-hand side of the above formula is the one used in the text.

## APPENDIX B: COMPUTATION OF THE BORDERED INFORMATION MATRIX OF FIML AND ITS INVERSE

The second-order differential of the log likelihood function is

$$\begin{aligned} d^2 \log L = & \frac{1}{2} \sum_i m_i \text{tr} \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} d\Sigma_i - NT \text{tr}(\Theta' L)^{-1} (d\Theta)' L (\Theta' L)^{-1} L' d\Theta \\ & - \frac{1}{2} \text{tr} \sum_i (d\Theta)' Z' M_i Z d\Theta \Sigma_i^{-1} + \frac{1}{2} \text{tr} \sum_i (d\Theta)' Z' M_i Z \Theta \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} \\ & - \frac{1}{2} \text{tr} \sum_i (d\Theta)' Z' M_i Z d\Theta \Sigma_i^{-1} + \frac{1}{2} \text{tr} \sum_i \Theta' Z' M_i Z d\Theta \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} \end{aligned}$$

$$\begin{aligned}
 & + \frac{1}{2} \operatorname{tr} \sum_i (d\Theta) Z' M_i Z \Theta \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} + \frac{1}{2} \operatorname{tr} \sum_i \Theta' Z' M_i Z d\Theta \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} \\
 & - \frac{1}{2} \operatorname{tr} \sum_i \Theta' Z' M_i Z \Theta \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} \\
 & - \frac{1}{2} \operatorname{tr} \sum_i \Theta' Z' M_i Z \Theta \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1}.
 \end{aligned} \tag{B.1}$$

Noting that

$$E(Z' M_i Z) = \begin{bmatrix} \Pi' \\ I \end{bmatrix} X' M_i X [\Pi I] + m_i L \Omega_i L',$$

$$E(Z' M_i Z \Theta) = m_i L \Omega_i L' \Theta,$$

$$E(\Theta' Z' M_i Z \Theta) = m_i \Theta' L \Omega_i L' \Theta = \Sigma_i,$$

and that

$$E(Z' M_i Z \Theta) \Sigma_i^{-1} = m_i L \Omega_i L' \Theta \Sigma_i^{-1} = m_i L (\Theta' L)^{-1},$$

and using the notation:

$$D_i = \begin{bmatrix} \Pi' \\ I \end{bmatrix} X' M_i X [\Pi I],$$

we can write:

$$\begin{aligned}
 -E(d^2 \log L) = & -\frac{1}{2} \sum_i m_i \operatorname{tr} \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} d\Sigma_i + NT \operatorname{tr} (\Theta' L)^{-1} (d\Theta)' L (\Theta' L)^{-1} d\Theta \\
 & + \frac{1}{2} \operatorname{tr} \sum_i (d\Theta)' (D_i + m_i L \Omega_i L') d\Theta \Sigma_i^{-1} - \frac{1}{2} \operatorname{tr} \sum_i m_i (d\Theta)' L (\Theta' L)^{-1} d\Sigma_i \Sigma_i^{-1} \\
 & + \frac{1}{2} \operatorname{tr} \sum_i (d\Theta)' (D_i + m_i L \Omega_i L') d\Theta \Sigma_i^{-1} - \frac{1}{2} \operatorname{tr} \sum_i m_i (L' \Theta) d\Theta \Sigma_i^{-1} d\Sigma_i \\
 & - \frac{1}{2} \operatorname{tr} \sum_i m_i (d\Theta)' L (\Theta' L)^{-1} d\Sigma_i \Sigma_i^{-1} - \frac{1}{2} \operatorname{tr} \sum_i m_i (L' \Theta)^{-1} d\Theta \Sigma_i^{-1} d\Sigma_i \\
 & + \frac{1}{2} \operatorname{tr} \sum_i m_i (d\Sigma_i) \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1} + \frac{1}{2} \operatorname{tr} \sum_i m_i d\Sigma_i \Sigma_i^{-1} d\Sigma_i \Sigma_i^{-1}.
 \end{aligned}$$

Thus, the information matrix  $\Psi$  is the symmetric matrix  $B$ , given on p. 241.

Let us give the blocks of the above  $\Psi$  the obvious notations  $\Psi_{ij}$ , where  $i, j = 1, 2, 3, 4$ . Now we have to take the limit of  $\eta' \Psi \eta$  where

$$\eta = \operatorname{diag} \left[ \frac{1}{\sqrt{NT}} I \quad \frac{1}{\sqrt{NT}} I \quad \frac{1}{\sqrt{N}} I \quad \frac{1}{\sqrt{T}} I \right], \quad \text{partitioned appropriately.}$$

Noting the following limits:

$$\begin{aligned}
\frac{1}{NT} X' M_4 X &= \frac{1}{NT} X' Q X \rightarrow S, \\
\Sigma_i^{-1} &\rightarrow 0, \quad i = 1, 2, 3 \\
T \Sigma_1^{-1} &\rightarrow \Sigma_\mu^{-1}, \\
N \Sigma_2^{-1} &\rightarrow \Sigma_\lambda^{-1}, \\
T \Sigma_3^{-1} &\rightarrow (\Sigma_\mu + \Sigma_\lambda)^{-1}, \quad \text{assuming } \frac{N}{T} \rightarrow 1 \\
N \Sigma_3^{-1} &\rightarrow (\Sigma_\mu + \Sigma_\lambda)^{-1}, \quad \text{assuming } \frac{N}{T} \rightarrow 1 \\
\frac{1}{T} \Omega_1 &\rightarrow \Omega_\mu, \\
\frac{1}{N} \Omega_2 &\rightarrow \Omega_\lambda, \\
\frac{1}{N} \Omega_3 &\rightarrow \Omega_\mu + \Omega_\lambda, \quad \text{assuming } \frac{N}{T} \rightarrow 1 \\
\frac{1}{T} \Omega_3 &\rightarrow \Omega_\mu + \Omega_\lambda, \quad \text{assuming } \frac{N}{T} \rightarrow 1
\end{aligned} \tag{B.3}$$

we find that:

$$\frac{1}{NT} \Psi_{11} \rightarrow [(L'\theta)^{-1} L' \otimes L(\theta' L)^{-1}] P + \Sigma_\nu^{-1} \otimes (D + L\Omega_\nu L')$$

$$\text{where } D = \begin{bmatrix} \Pi' \\ I \end{bmatrix} S[\Pi I].$$

Now, by writing

$$\begin{aligned}
[(L'\theta)^{-1} L' \otimes L(\theta' L)^{-1}] P &= [I \otimes L(\theta' L)^{-1}] [(L'\theta)^{-1} L' \otimes I] P \\
&= [I \otimes L(\theta' L)^{-1}] P [I \otimes (L'\theta)^{-1} L' \otimes I] \\
&= [\Sigma_\nu^{-1} \otimes L(\theta' L)^{-1}] [\Sigma_\nu \otimes I] P [\Sigma_\nu \otimes I] \\
&\quad \times [\Sigma_\nu^{-1} \otimes (L'\theta)^{-1} L'] \\
&= [\Sigma_\nu^{-1} \otimes L(\theta' L)^{-1}] [\Sigma_\nu \otimes \Sigma_\nu] P [\Sigma_\nu^{-1} \otimes (L'\theta)^{-1} L']
\end{aligned}$$

and

$$\begin{aligned}
\Sigma_\nu^{-1} \otimes L\Omega_\nu L' &= \Sigma_\nu^{-1} \otimes L(\theta' L)^{-1} \Sigma_\nu (L'\theta)^{-1} L' \\
&= [\Sigma_\nu^{-1} \otimes L(\theta' L)^{-1}] [\Sigma_\nu \otimes \Sigma_\nu] [\Sigma_\nu^{-1} \otimes (L'\theta)^{-1} L']
\end{aligned}$$

and calling

$$H_{12} = -[\Sigma_\nu^{-1} \otimes L(\theta' L)^{-1}], \tag{B.4}$$

we get

$$\frac{1}{NT} \Psi_{11} \rightarrow \Sigma_\nu^{-1} \otimes D + H_{12} (2\Sigma_\nu \otimes \Sigma_\nu) \frac{I + P}{2} H_{12}' \equiv H_{11}. \tag{B.5}$$

Using the limits given by (B.3), it is straightforward that

$$\frac{1}{NT} \Psi_{22} \rightarrow \frac{1}{2} \Sigma_\nu^{-1} \otimes \Sigma_\nu^{-1} \equiv H_{22}, \quad (\text{B.6})$$

$$\frac{1}{N} \Psi_{33} \rightarrow \frac{1}{2} \Sigma_\mu^{-1} \otimes \Sigma_\mu^{-1} \equiv H_{33}, \quad (\text{B.7})$$

$$\frac{1}{T} \Psi_{44} \rightarrow \frac{1}{2} \Sigma_\lambda^{-1} \otimes \Sigma_\lambda^{-1} \equiv H_{44}, \quad (\text{B.8})$$

and that all the remaining blocks tend to zero when divided by the appropriate factor.

Therefore, ordering the parameters as  $\text{vec } \Theta$ ,  $\text{vec } \Sigma_\nu$ ,  $\text{vec } \Sigma_\mu$ , and  $\text{vec } \Sigma_\lambda$ , the information matrix (in limit), bordered by the linear restrictions on  $\Theta$  and the symmetry conditions on the covariance matrices, is:

$$H = \left[ \begin{array}{cccc|cccc} H_{11} & H_{12} & 0 & 0 & R' & 0 & 0 & 0 \\ H_{12}' & H_{22} & 0 & 0 & 0 & C' & 0 & 0 \\ 0 & 0 & H_{33} & 0 & 0 & 0 & C' & 0 \\ 0 & 0 & 0 & H_{44} & 0 & 0 & 0 & C' \\ \hline R & 0 & 0 & 0 & & & & \\ 0 & C & 0 & 0 & & & & \\ 0 & 0 & C & 0 & & & & \\ 0 & 0 & 0 & C & & & & \end{array} \right].$$

Let us denote by  $H^{ij}$  the corresponding block of the inverse of  $H$ . Then, as in the case of the reduced form, we have:

$$H^{33} = \frac{1}{2} (I + P) (2\Sigma_\mu \otimes \Sigma_\mu) \frac{1}{2} (I + P), \quad (\text{B.9})$$

$$H^{44} = \frac{1}{2} (I + P) (2\Sigma_\lambda \otimes \Sigma_\lambda) \frac{1}{2} (I + P). \quad (\text{B.10})$$

For the first four blocks of  $H^{-1}$ , we notice that they are equal to the first four blocks of the inverse of the following matrix  $A$ :

$$A = \left[ \begin{array}{cc|cc} H_{11} & H_{12} & R' & 0 \\ H_{12}' & H_{22} & 0 & C' \\ \hline R & 0 & 0 & 0 \\ 0 & C & 0 & 0 \end{array} \right] = \begin{bmatrix} A_1 & G' \\ G & 0 \end{bmatrix}.$$

Construct the following matrix  $F_*$

$$F_* = \begin{bmatrix} F & 0 \\ 0 & \tilde{F} \end{bmatrix},$$

where  $F$  is a matrix of orthonormal vectors such that  $FF' = I - R'(RR')^{-1}R$  and  $\tilde{F}$  is a matrix of orthonormal vectors such that  $\tilde{F}\tilde{F}' = I - C'C = \frac{1}{2}(I + P)$ . It is easily checked that  $F_*F'_* = I - G'(GG')^{-1}G$ . Therefore, the inverse corresponding to  $A_1$ , say  $A^1$ , is given by

$$A^1 = F_*(F'_*A_1F_*)^{-1}F'_*.$$

Let us now compute the above matrix. First, we write:

$$F'_*A_1F_* = \begin{bmatrix} F'H_{11}F & F'H_{12}\tilde{F} \\ \tilde{F}'H'_{12}F & \tilde{F}'H_{22}\tilde{F} \end{bmatrix},$$

and notice that the block  $\tilde{F}'H_{22}\tilde{F}$  is nonsingular. Therefore, using the inversion rule for partitioned matrices and upon pre-multiplication by  $F_*$  and post-multiplication by  $F'_*$ , we obtain:

$$H^{11} = F\tilde{S}^{-1}F',$$

$$H^{12} = F\tilde{S}^{-1}F'H_{12}\tilde{F}(\tilde{F}'H_{22}\tilde{F})^{-1}\tilde{F},$$

$$H^{22} = \tilde{F}(\tilde{F}'H_{22}\tilde{F})^{-1}[I + \tilde{F}'H'_{12}F\tilde{S}^{-1}F'H_{12}\tilde{F}(\tilde{F}'H_{22}\tilde{F})^{-1}]\tilde{F}',$$

where

$$\tilde{S} = F'H_{11}F - F'H_{12}\tilde{F}(\tilde{F}'H_{22}\tilde{F})^{-1}\tilde{F}'H'_{12}F,$$

$$= F'H_{11}F - F'H_{12}\tilde{F}\tilde{F}'H_{22}^{-1}\tilde{F}\tilde{F}'H'_{12}F,$$

$$= F'(\Sigma_v^{-1} \otimes D)F.$$

Therefore, we can simplify the above results, to obtain:

$$H^{11} = F[F'(\Sigma_v^{-1} \otimes D)F]^{-1}F', \quad (\text{B.11})$$

$$H^{12} = F[F'(\Sigma_v^{-1} \otimes D)F]^{-1}F'H_{12}\tilde{F}\tilde{F}'H_{22}^{-1}\tilde{F}\tilde{F}', \quad (\text{B.12})$$

$$\begin{aligned} H^{22} &= \tilde{F}(\tilde{F}'H_{22}\tilde{F})^{-1}(\tilde{F}' + \tilde{F}'H'_{12}F[F'(\Sigma_v^{-1} \otimes D)F]^{-1}\tilde{F}'H_{12}\tilde{F}\tilde{F}'H_{22}\tilde{F}\tilde{F}'), \\ &= \tilde{F}\tilde{F}'H_{22}^{-1}\tilde{F}\tilde{F}' + \tilde{F}\tilde{F}'H_{22}^{-1}\tilde{F}\tilde{F}'H'_{12}F[F'(\Sigma_v^{-1} \otimes D)F]^{-1}F'H_{12}\tilde{F}\tilde{F}'H_{22}^{-1}\tilde{F}\tilde{F}', \\ &= \tilde{F}\tilde{F}'H_{22}^{-1}\tilde{F}\tilde{F}' + \tilde{F}\tilde{F}'H_{22}^{-1}H'_{12}F[F'(\Sigma_v^{-1} \otimes D)F]^{-1}F'H_{12}H_{22}^{-1}\tilde{F}\tilde{F}', \\ &= \frac{1}{2}(I + P)(2\Sigma_v \otimes \Sigma_v) \frac{1}{2}(I + P) + \frac{1}{2}(I + P)(2I \otimes \Sigma_v(L'\Theta)^{-1}L')F \\ &\quad \times [F'(\Sigma_v^{-1} \otimes D)F]^{-1}F'(2I \otimes L(\Theta'L)^{-1}\Sigma_v) \frac{1}{2}(I + P). \end{aligned} \quad (\text{B.13})$$